NONPARAMETRIC ESTIMATION OF BOUNDARIES*

E. Carlstein and C. Krishnamoorthy

University of North Carolina, Chapel Hill

ABSTRACT

A data-set consists of independent observations taken at the nodes of a grid. An unknown *boundary* partitions the grid into two regions: All the observations coming from a particular region share a common distribution, but the distributions are different for the two different regions. These two distributions are entirely unknown, and they need not differ in their means, medians, or other measure of "level." The grid is of arbitrary dimension, and its mesh need not be squares. Our objective is to estimate the boundary, using only the observed data.

A class of nonparametric estimators is proposed. We obtain strong consistency for these estimators (including rates of convergence and a bound on the error probability). The only assumptions needed are some mild and easily verifiable regularity conditions on boundaries.

The boundary-estimation problem has applications in diverse fields, including: quality control, epidemiology, forestry, marine science, meteorology, and geology. Our method provides (as special cases) nonparametric estimators for the following situations: the change-point problem; the epidemic-change model; templates; linear bisection of the plane; Lipschitz boundaries. Each of these applications is explicitly illustrated.

* Research supported by National Science Foundation Grant #DMS-8701201, and by a U.N.C. Junior Faculty Development Award.

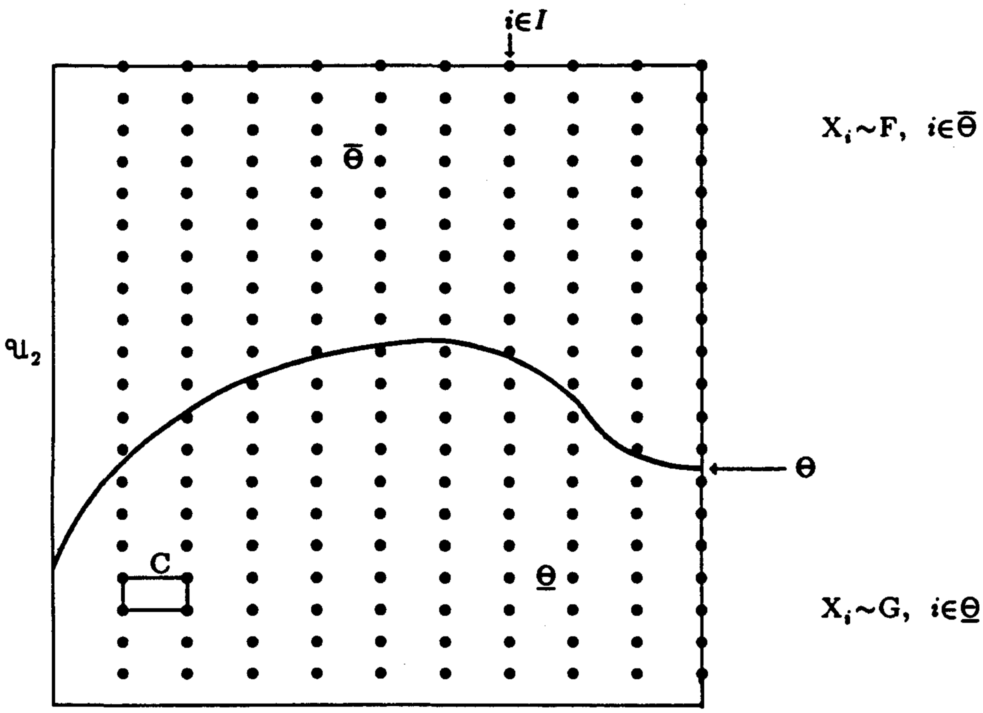
KEY WORDS: change-point, Cramér-von Mises, empirical c.d.f., epidemic-change, grid, Kolmogorov-Smirnov, Lipschitz, partition, template.

1. INTRODUCTION

1.1 The Statistical Problem. We observe a collection of independent r.v.s $\{X_i\}$, indexed by nodes i of a finite d-dimensional grid. Without loss of generality, the grid is taken to be in the d-dimensional unit cube $\mathfrak{U}_d:=[0,\ 1]^d$. The unknown *boundary* Θ is simply a $(d-1)$-dimensional surface that partitions $\mathfrak{U}_d$ into two regions, $\overline{\Theta}$ and $\underline{\Theta}$. All observations X_i made at nodes $i\in\overline{\Theta}$ are from distribution F, while all observations X_i made at nodes $i\in\underline{\Theta}$ are from distribution G. The objective is to estimate the unknown boundary Θ, using only the observed data $\{X_i\}$. [Figure 1 illustrates the set-up in the case d=2.]

FIGURE 1

The 2-Dimensional Case: $\mathfrak{U}_2$ with $n_1=10$, $n_2=20$, $|I|=200$.



1.2 Distributional Assumptions. The distributions F and G are entirely unknown. We will not assume any knowledge of the functional forms or parametric families of F and G. No regularity conditions (e.g., continuity, discreteness) will be imposed on F and G. No prior information is needed regarding *how* F and G differ (e.g., they need not differ in their means, medians, or other measure of "level"). *The only distributional assumption is that* F$\neq$G. Our fully nonparametric method is needed when the user has insufficient prior knowledge of the underlying distributions, or when the user wants a robust corroborator for results from a parametric analysis.

1.3 Grid Assumptions. The grid is generated by divisions along each coordinate axis in $\mathfrak{U}_d$ [see Figure 1]. Along the j^{th} axis ($1\leq j\leq d$), there are n_j divisions which are equally spaced at $1/n_j$, $2/n_j$, ..., n_j/n_j. Observations are made at the resulting *grid nodes* $i:=(i_1/n_1, i_2/n_2, \ldots, i_d/n_d)\in\mathfrak{U}_d$, where $i_j\in\{1, 2, \ldots, n_j\}$. Thus the grid mesh need not be squares. A rectangular mesh allows for different sampling designs along the different dimensions; this in turn may reflect differing sampling *costs* in the different dimensions [see Examples]. The collection of all nodes i is denoted by I, and the total number of observations is $|I|:=\prod_{j=1}^{d} n_j$.

1.4 Boundaries. The notion of a *boundary* in $\mathfrak{U}_d$ is formulated in a set-theoretic way: the unknown boundary Θ is identified with the corresponding partition ($\overline{\Theta}$, $\underline{\Theta}$) of $\mathfrak{U}_d$. This general formulation is free of the dimension d, and allows the maximum flexibility for treating a wide variety of specific situations [see Section 2]. Note that $\overline{\Theta}$ need not be a convex set [Examples 2.b, 2.c], and $\overline{\Theta}$ need not even be a connected set [Example 1.b].

The sample-based *estimate* of Θ will be selected from a finite collection $\mathfrak{T}$ of *candidate boundaries*, with generic element T. Again, each candidate T is identified with its corresponding partition ($\overline{T}$, $\underline{T}$) of $\mathfrak{U}_d$. The total number of candidates considered is $|\mathfrak{T}|:=\#\{T\in\mathfrak{T}\}$.

In any set $A\subseteq\mathfrak{U}_d$, the number of observations (i.e., grid nodes) is $|A|:=\#\{i\in A\}$.

1.5 Outline of the Paper. The notions discussed in this Introduction will be illustrated by specific examples in Section 2. The proposed nonparametric boundary estimator is described in Section 3; we also compare our estimator to the related (but less general) estimators in the literature. Theoretical properties of our estimator are presented in Section 4, but the proofs of these are deferred to the Appendix. In Section 5, we explicitly apply our nonparametric method -- and the theoretical results -- to the examples from Section 2.

2. EXAMPLES

2.1 The 1-Dimensional Case. In the case $d=1$, it is natural to think of $\mathcal{U}_d$ as a "time" axis; then I indexes observations at equally spaced intervals of time.

EXAMPLE 1.*a*: The Change-Point Problem. The boundary Θ is simply an arbitrary real number $\theta \in (0, 1)$, inducing the partition: $\overline{\Theta} := [0, \theta)$, $\underline{\Theta} := [\theta, 1]$. Estimation of the "change-point" Θ has been extensively studied in the literature [see Shaban (1980) for an annotated bibliography]; most of this other work assumes either parametric knowledge of F and G, or assumes that F and G differ in a known way (e.g., by a shift in "level"). In the quality-control setting, Θ demarcates a change from an "in-control" production process to an "out-of-control" production process. Since we make no assumptions about G, our method will identify the onset of *any* type of disorder in the distribution of the output.

The candidate boundaries $T \in \mathcal{T}$ are essentially all the times at which observations were made, i.e., all $t \in \{2/n_1, 3/n_1, \ldots, (n_1-1)/n_1\}$, with $\overline{T} := [0, t)$, $\underline{T} := [t, 1]$. It is natural to "anchor" the Ts to the grid nodes in I, because in practice one cannot hope to get better resolution from an *estimated* boundary than whatever degree of resolution is available from the *data*-nodes I.

EXAMPLE 1.*b*: The Epidemic-Change Model. The boundary Θ consists of two points $\{\theta_1, \theta_2\}$, with $0<\theta_1<\theta_2<1$, inducing the partition: $\overline{\Theta} := [0, \theta_1) \cup [\theta_2, 1]$, $\underline{\Theta} := [\theta_1, \theta_2)$. In this model, θ_1 represents a change to the epidemic distribution G, and θ_2 represents a return to the pre-epidemic distribution F. Again, previous work on the epidemic-change model requires parametric knowledge of F and G, and/or assumes a shift in "level" [see Siegmund (1986) and Bhattacharya & Brockwell (1976)]. Note that $\overline{\Theta}$ is not even a connected set in this example.

The candidate boundaries $T \in \mathcal{T}$ are all pairs $\{t_1, t_2\}$ anchored in I, with: $t_1<t_2$, $t_1 \in \{1/n_1, 2/n_1, \ldots, (n_1-2)/n_1\}$, $t_2 \in \{2/n_1, 3/n_1, \ldots, (n_1-1)/n_1\}$, $\overline{T} := [0, t_1) \cup [t_2, 1]$, $\underline{T} := [t_1, t_2)$.

2.2 The 2-Dimensional Case. In the case $d=2$, it is natural to think of $\mathcal{U}_d$ as a geographic area; then I indexes observations which are regularly spaced in the East-West direction and which are also regularly spaced in the North-South direction.

Consider the following application from forestry: The observations $\{X_i: i \in I\}$ represent heights of trees, where F is the distribution for a healthy stand and G is the distribution for a diseased stand. If the disease kills very young and very old trees, then F and G may share the same "level," but may differ in terms of "dispersion." The point here is that our estimator will identify the boundary between "healthy" and "diseased" *without any prior knowledge regarding the effect of the disease on the distribution of heights.* If the disease spreads radially through the population, then the boundaries considered should be circular or elliptical templates [see Example 2.b].

Marine scientists often rely on voyages of commercial vessels in order to obtain data. When a trans-Atlantic voyage is being made, it is relatively inexpensive to record observations at a large number (n_1) of closely-spaced intervals in the East-West direction. But it is very expensive to extend the grid in the North-South direction (i.e., to increase n_2), since this entails a whole new trans-Atlantic voyage. So it is important in practice to allow for different sampling designs in the different dimensions. Our method does allow the n_js to differ, and our theoretical analysis [Section 5] shows how the n_js affect the rate of convergence for our estimator.

EXAMPLE 2.a: Linear Bisection. The boundary Θ is an arbitrary straight line-segment connecting endpoints on two distinct edges of $\mathcal{U}_2$; this boundary induces a partition $(\overline{\Theta}, \underline{\Theta})$ in an obvious way.

The candidate boundaries $T \in \mathcal{T}$ are all straight line-segments connecting endpoints on two distinct edges of $\mathcal{U}_2$ -- but the endpoints must be anchored to the grid: An endpoint on an East-West edge must have coordinate i_1/n_1, $i_1 \in \{2, 3, \ldots, n_1-1\}$, while an endpoint on a North-South edge must have coordinate i_2/n_2, $i_2 \in \{2, 3, \ldots, n_2-1\}$.

EXAMPLE 2.b: Templates. A *template* is a boundary in $\mathcal{U}_2$ which can be perturbed via a finite number of "parameters;" these "parameters" may allow for translation, rotation, elongation, etc. Circles, ellipses, and polygons can be handled as templates.

Consider, for example, an arbitrary rectangular template Θ whose edges are parallel to the edges of $\mathcal{U}_2$, and whose vertices are all in $(0, 1)^2$. The interior of the rectangle is $\underline{\Theta}$, and the remainder of $\mathcal{U}_2$ is $\overline{\Theta}$. Note that $\overline{\Theta}$ is not even a convex set.

The candidate boundaries $T \in \mathcal{T}$ are all rectangles whose edges are parallel to the edges of $\mathcal{U}_2$, and whose vertices are anchored to grid nodes: The East-West coordinate of a vertex must be in $\{1/n_1, 2/n_1, \ldots, (n_1-1)/n_1\}$, and the North-South coordinate must be in $\{1/n_2, 2/n_2, \ldots, (n_2-1)/n_2\}$. The region $\underline{T}$ is the interior of the rectangle together with its edges.

EXAMPLE 2.c: Lipschitz Boundaries. To define a Lipschitz boundary, we identify the lower edge of $\mathcal{U}_2$ as the z-axis, and the left-hand edge of $\mathcal{U}_2$ as the y-axis. A *Lipschitz boundary* Θ is the curve in $\mathcal{U}_2$ corresponding to a function $y_\Theta(\cdot)$: $[0, 1] \mapsto (0, 1)$ satisfying: $|y_\Theta(z)-y_\Theta(z')| \leq c_0|z-z'| \ \forall \ z,z' \in [0, 1]$. The constant c_0 simply controls the slope of the boundary. The points $(y, z) \in \mathcal{U}_2$ with $y > y_\Theta(z)$ comprise the region $\overline{\Theta}$; the remainder of $\mathcal{U}_2$ is $\underline{\Theta}$.

The class of Lipschitz boundaries is extremely rich. Note that these boundaries are not readily expressible as templates, and that $\overline{\Theta}$ and $\underline{\Theta}$ need not be convex sets.

The candidate boundaries $T \in \mathcal{T}$ correspond to piecewise linear functions $y_T(\cdot)$: $[0, 1] \mapsto (0, 1)$ which are anchored to the grid in the following way: For each $z \in \{0/n_1, 1/n_1, \ldots, n_1/n_1\}$, the associated value of $y_T(z)$ is in $\{1/n_2, 2/n_2, \ldots, (n_2-1)/n_2\}$; at intermediate values of z, the function $y_T(z)$ is defined by linear interpolation. We restrict $\mathcal{T}$ to those boundaries T for which $|y_T(z)-y_T(z-\frac{1}{n_1})| \leq c/n_1 \ \ \forall z \in \{1/n_1, 2/n_1, \ldots, n_1/n_1\}$, where $c := 3c_0+1$. This Lipschitz-type restriction controls the slope of the candidate boundaries. For this Example, it is convenient to have $n_2 \geq n_1$. The region $\overline{T}$ is defined analogously to $\overline{\Theta}$.

2.3 **The 3-Dimensional Case.** Boundary estimation in the case $d=3$ has natural applications to meteorology and geology. Planar bisection of $\mathcal{U}_3$ and templates in $\mathcal{U}_3$ can be handled analogously to Examples 2.a and 2.b above. In geology, the cost of extending the grid in the "depth" direction may again necessitate a sampling design with differing n_js.

3. THE BOUNDARY ESTIMATOR

3.1 **The Basic Idea.** Our main statistical tool for selecting an estimate $\hat{\Theta}$ from $\mathcal{T}$ is the *empirical cumulative distribution function (e.c.d.f.)*. For a candidate boundary $T \in \mathcal{T}$, compute the e.c.d.f.

$$\bar{h}_T(x) := \sum_{i \in \bar{T}} I\{X_i \leq x\}/|\bar{T}|,$$

which treats all observations from region $\bar{T}$ *as if* they were identically distributed; similarly compute

$$\underline{h}_T(x) := \sum_{i \in \underline{T}} I\{X_i \leq x\}/|\underline{T}|,$$

which treats all observations from region $\underline{T}$ *as if* they were identically distributed. The former e.c.d.f. is actually a sample estimate of the unknown mixture distribution

$$[|\bar{T} \cap \bar{\Theta}|F(x) + |\bar{T} \cap \underline{\Theta}|G(x)]/|\bar{T}|,$$

while the latter e.c.d.f. analogously estimates

$$[|\underline{T} \cap \bar{\Theta}|F(x) + |\underline{T} \cap \underline{\Theta}|G(x)]/|\underline{T}|.$$

Therefore, the *difference* between the two e.c.d.f.s can be approximated as follows:

$$|\bar{h}_T(x) - \underline{h}_T(x)| \approx |[|\bar{T} \cap \bar{\Theta}|/|\bar{T}|] - [|\underline{T} \cap \bar{\Theta}|/|\underline{T}|]| \cdot |F(x) - G(x)|.$$

Notice that this last expression can never exceed $|F(x)-G(x)|$ (because $(\bar{T} \cap \bar{\Theta}) \subseteq \bar{T}$ and $(\underline{T} \cap \bar{\Theta}) \subseteq \underline{T}$); moreover, its maximizing value $|F(x)-G(x)|$ is attained *precisely when* $T=\Theta$. This suggests a natural approach for estimating Θ: Choose as your estimator the candidate boundary which maximizes $|\bar{h}_T(x) - \underline{h}_T(x)|$ over all $T \in \mathcal{T}$. This basic idea will now be refined and generalized.

3.2 Definition of the Boundary Estimator. Rather than restricting attention to the difference $|\bar{h}_T(\cdot)-\underline{h}_T(\cdot)|$ at a *single specified* x-value, we instead consider the differences

$$d_i^T := |\bar{h}_T(X_i)-\underline{h}_T(X_i)|$$

for each $i \in I$. This allows the data to lead us toward the informative x-values, without prior knowledge of F and G.

Now combine these differences d_i^T using a general norming function $S(d_1^T, d_2^T, \ldots, d_{|I|}^T)$. The norm $S(\cdot)$ must satisfy certain simple conditions [see Section 4], but special cases include:

the *Kolmogorov-Smirnov* norm $S_{KS}(d_1, d_2, \ldots, d_N) := sup_{1 \le i \le N}\{d_i\}$;

the *Cramér-von Mises* norm $S_{Cv}(d_1, d_2, \ldots, d_N) := (\sum_{1 \le i \le N} d_i^2/N)^{1/2}$;

the *arithmetic-mean* norm $S_{am}(d_1, d_2, \ldots, d_N) := \sum_{1 \le i \le N} d_i/N$.

Finally, we must standardize to account for the inherent instability in the e.c.d.f. Suppose $\bar{h}_{T_o}(\cdot)$ (say) is based on a very small amount of data (i.e., $|\bar{T}_0|$ is very small relative to $|I|$); then $d_i^{T_o}$ is unstable. Thus $d_i^{T_o}$ may be "large" (as compared to other d_i^Ts) merely due to random variability. We should *downweight* this particular candidate T_0 in our search through $\mathfrak{T}$ for a maximizer. This downweighting is accomplished by the multiplicative factor $\lambda(\bar{T}_0)$, where $\lambda(\cdot)$ is Lebesgue measure over $\mathfrak{U}_d$. Since $\lambda(\cdot)$ is like "area," this standardizing factor is essentially the proportion of data in $\bar{T}_0$, i.e., $|\bar{T}_0|/|I|$.

The nonparametric boundary estimator $\hat{\Theta}$ is defined as the candidate boundary in $\mathfrak{T}$ which maximizes the criterion function

$$D(T) := \lambda(\bar{T})\lambda(\underline{T})\, S(d_1^T, d_2^T, \ldots, d_{|I|}^T)$$

over all $T \in \mathfrak{T}$. Formally,

$$\hat{\Theta} := argmax_{T \in \mathfrak{T}}\, D(T).$$

Observe that $\hat{\Theta}$ is calculated solely from the data at hand. Theoretical properties of $\hat{\Theta}$ are presented in Section 4.

3.3 Related Methods. Our proposed approach is related to several other methods that have recently been studied in the literature. However, each of these other methods suffers from at least one of the following limitations:

(a) Restrictions on F and G must be made; e.g., F and G are both assumed to be Gaussian, or F and G are both assumed to be discrete with finite supports.

(b) Only the 1-dimensional single-change-point problem is considered [our Example 1.a].

(c) The norm $S(\cdot)$ must be specifically of the form $S_{KS}(\cdot)$ or $S_{Cv}(\cdot)$.

For example, Deshayes & Picard (1981) propose a hypothesis testing procedure in the 1-dimensional single-change-point scenario, based specifically on $S_{KS}(\cdot)$; note that they are testing the null hypothesis of "no change" vs the alternative "*some* change occurs," rather than testing or estimating *the specific location of the boundary.* Picard (1985) also tests for the existence of a single change-point, under the assumption that F and G are *both Gaussian with common mean*; she uses $S_{KS}(\cdot)$ and $S_{Cv}(\cdot)$. Darkhovskii & Brodskii (1980) consider the 1-dimensional single-change-point problem under the assumptions that F and G are both *discrete* with *finite supports*, and that Θ is in a *known* interval bounded away from the endpoints of (0, 1); they use $S_{Cv}(\cdot)$. Darkhovskii (1984) & (1986) deals with the case $d=1$, again assuming that F and G are both discrete with finite supports, and again using $S_{Cv}(\cdot)$. Darkhovskii (1985) also considers the case $d=1$, but uses $S_{KS}(\cdot)$. Brodskii & Darkhovskii (1986) again assume that F and G are both discrete with finite supports; and, they only consider the norm $S_{Cv}(\cdot)$. Asatryan & Safaryan (1986) study the 1-dimensional single-change-point problem; they assume that Θ is in a known interval bounded away from the endpoints of (0, 1), and that F and G are both *continuous.* Csörgő & Horváth (1987) deal with the 1-dimensional single-change-point problem, assuming that F and G are both continuous; they use $S_{KS}(\cdot)$. Carlstein (1988) considers only the 1-dimensional single-change-point problem.

Our proposed method entirely avoids limitations of types (a), (b), and (c): *No restrictions* (e.g., parametric conditions or regularity conditions) on F or G are needed, because we rely exclusively on the e.c.d.f. Our *set-theoretic* formulation of boundaries is natural for index-grids of *any dimension*, and it allows for a wide variety of examples [see Section 2]. We abstract the important *properties* of the norm $S(\cdot)$, and then allow *any* choice of norm with those properties; $S_{KS}(\cdot)$ and $S_{Cv}(\cdot)$ are just two special cases handled within our unified formulation.

4. THEORETICAL PROPERTIES OF THE BOUNDARY ESTIMATOR

4.1 Measuring the "Distance" Between Boundaries. In order to assess the performance of our boundary estimator $\hat{\Theta}$, we must first quantify the notion of "distance" between two boundaries (say, T and Θ). Our "distance" measure is

$$\partial(T, \Theta) := \min\{\lambda(\bar{T} \circ \bar{\Theta}), \lambda(\underline{T} \circ \bar{\Theta})\},$$

where $\circ$ denotes set-theoretic symmetric difference, i.e., $(A \circ B) := (A \cap B^c) \cup (A^c \cap B)$. Intuitively, $\lambda(\bar{T} \circ \bar{\Theta})$ represents the "area" that is misclassified by $\bar{T}$ as an estimator of $\bar{\Theta}$. For a given boundary T, the *a priori* labelling of the two induced regions as $(\bar{T}, \underline{T})$ [rather than $(\underline{T}, \bar{T})$] can be arbitrary [e.g., Example 2.a]. Therefore, a candidate boundary T is considered "close" to Θ if *either* $\bar{T}$ *or* $\underline{T}$ is nearly the same region of $\mathfrak{U}_d$ as $\bar{\Theta}$.

The function $\partial(\cdot\,,\cdot)$ has the following desirable properties of a "distance."

***PROPOSITION* 1:** The function $\partial(\cdot\,,\cdot)$ is a *pseudometric*. That is, it satisfies:

(1.*a*) [Non-negativity] $\partial(T, \Theta) \geq 0$.

(1.*b*) [Identity] $\partial(\Theta, \Theta) = 0$.

(1.*c*) [Symmetry] $\partial(T, \Theta) = \partial(\Theta, T)$.

(1.*d*) [Triangle Inequality] $\partial(T, \Theta) \leq \partial(T, T') + \partial(T', \Theta)$.

Properties 1.a, 1.b, and 1.c are obvious from the definition; a proof of property 1.d is in Appendix A.1.

When we discuss "consistency" and probability of "error" for $\hat{\Theta}$ as an estimator of Θ, it will always be in the sense of ∂-distance.

4.2 The Norm $S(\cdot)$. There are some constraints on the choice of $S(\cdot)$. The following conditions are intuitively reasonable, and they enable us to simultaneously handle a whole *class* of boundary estimators $\hat{\Theta}$.

DEFINITION: A function $S(\cdot)$: $\mathbf{R}^N_+ \mapsto \mathbf{R}^1_+$ is a *mean-dominant norm* if it satisfies:

(*D.a*) [Symmetry] $S(\cdot)$ is symmetric in its N arguments.

(*D.b*) [Homogeneity] $S(\alpha d_1, \alpha d_2, \ldots, \alpha d_N) = \alpha S(d_1, d_2, \ldots, d_N)$ whenever $\alpha \geq 0$.

(*D.c*) [Triangle Inequality] $S(d_1+d_1', d_2+d_2', \ldots, d_N+d_N') \leq S(d_1, d_2, \ldots, d_N) + S(d_1', d_2', \ldots, d_N')$.

(*D.d*) [Identity] $S(1, 1, \ldots, 1) = 1$.

(*D.e*) [Monotonicity] $S(d_1, d_2, \ldots, d_N) \leq S(d_1', d_2', \ldots, d_N')$ whenever $d_i \leq d_i'$ $\forall i$.

(*D.f*) [Mean Dominance] $S(d_1, d_2, \ldots, d_N) \geq \sum_{1\leq i\leq N} d_i/N$.

It is straightforward to check that:

PROPOSITION 2: The functions $S_{KS}(\cdot)$, $S_{Cv}(\cdot)$, and $S_{am}(\cdot)$ are *mean-dominant norms.*

4.3 Asymptotic Results. In order to study the asymptotic properties of our method, we will let the number of grid-nodes increase: $|I| \rightarrow \infty$. Since the candidate boundaries and the estimator depend on the particular grid, we henceforth equip $\mathfrak{T}$ and $\hat{\Theta}$ with explicit subscripts: $\mathfrak{T}_I$ and $\hat{\Theta}_I$. Similarly, the number of observations in $A \subseteq \mathfrak{U}_d$ is now $|A|_I$.

We assume that: $F \neq G$; $\hat{\Theta}_I$ is based on a *mean-dominant norm*; the boundaries satisfy *regularity conditions* R.1–R.4 (described in Section 4.4). The main theoretical results are:

THEOREM 1: [Strong Consistency]

$$|I|^{\delta} \cdot \partial(\Theta, \hat{\Theta}_I) \rightarrow 0 \quad \text{as } |I| \rightarrow \infty, \text{ with probability 1.}$$

THEOREM 2: [Bound on Error Probability]

$$P\{\partial(\Theta, \hat{\Theta}_I) > \varepsilon\} \leq c_1 \cdot |\mathfrak{T}_I| \cdot exp\{-c_2 \cdot \varepsilon^2 \cdot |I|\} \quad \text{for } |I| \text{ sufficiently large,}$$

where $c_1 > 0$ and $c_2 > 0$ are constants.

Proofs of these results are in Appendix A.2. The "rate" of convergence obtained in Theorem 1 depends upon $\delta \geq 0$. Constraints on δ will follow from the regularity conditions discussed below. In Section 5 we will see the actual rates that can be obtained in particular applications. Theorem 2 says that the probability of error decreases exponentially as a function of sample size, but that this effect is counterbalanced by the number of candidate boundaries considered. The precise nature of this trade-off between $|I|$ and $|\mathfrak{T}_I|$ is discussed below in Section 4.4.

4.4 Regularity Conditions on Boundaries. Although Theorems 1 & 2 require no *distributional* assumptions, they do assume certain *set-theoretic* regularity conditions on the boundaries. It will be seen that these regularity conditions are intuitively natural, and that they are simple to check in specific applications [Section 5]. Moreover, these regularity conditions are essentially *constraints on* $\mathfrak{T}_I$; since $\mathfrak{T}_I$ is chosen at the discretion of the user, it is the user's prerogative to select $\mathfrak{T}_I$ in such a way that these constraints are satisfied.

REGULARITY CONDITION (R.1): [Non-trivial Partitions]

For each $T \in \mathfrak{T}_I$,

$$0<\lambda(\overline{T})<1 \text{ and } 0<|\overline{T}|_I/|I|<1.$$

Also, $0<\lambda(\overline{\Theta})<1$.

Condition R.1 prohibits consideration of trivial partitions.

REGULARITY CONDITION (R.2): [Richness of $\mathfrak{T}_I$]

For each I, $\exists\ T_I \in \mathfrak{T}_I$ such that the sequence $\{T_I\}$ satisfies:

$$|I|^{\delta} \cdot \partial(\Theta, T_I) \to 0 \text{ as } |I| \to \infty.$$

Condition R.2 requires $\mathfrak{T}_I$ to contain *some* candidate boundary T_I that "gets close" to the true Θ (at a rate corresponding to that desired in Theorem 1). If no such "ideal" candidate were even available, we could not possibly hope to statistically select an estimator $\hat{\Theta}_I$ from $\mathfrak{T}_I$ in such a way that the convergence of Theorem 1 holds. It is easy to satisfy R.2 when the candidate boundaries are anchored to the grid in a natural way [see Sections 2 & 5].

REGULARITY CONDITION (R.3): [Cardinality of $\mathfrak{T}_I$]

For each $\gamma>0$,

$$|\mathfrak{T}_I|\cdot exp\{-\gamma\cdot|I|^{1-2\delta}\} \to 0 \text{ as } |I|\to\infty.$$

Condition R.3 quantifies the balance between $|I|$ and $|\mathfrak{T}_I|$. Basically, the number of candidate boundaries must be substantially smaller than an exponential of the sample size. This constraint still allows for extremely rich collections $\mathfrak{T}_I$ [see Section 5]. In light of R.3, we can only obtain rates of convergence in Theorem 1 with $\delta<\frac{1}{2}$. Note that:

PROPOSITION 3:

Condition R.3 is satisfied for all $\delta\in[0, \frac{1}{2})$ whenever $|\mathfrak{T}_I|$ is of the order $|I|^\nu$, $\nu>0$.

In order to discuss the final regularity condition, we need the notion of *cells* in $\mathcal{U}_d$. Recall that the grid (described in Section 1.3) induces a "rectangular" partition of $\mathcal{U}_d$. A *cell* is simply one of these d-dimensional "rectangular" regions, including its edges and vertices [see Figure 1]. Thus there are $|I|$ cells in $\mathcal{U}_d$, and they are not strictly disjoint. A generic cell is denoted C, and the collection of all cells in $\mathcal{U}_d$ is denoted $\mathfrak{C}_I$. For an arbitrary set $A\subseteq\mathcal{U}_d$, the collection of *perimeter cells of A* is defined as:

$$\mathcal{P}_I(A):=\{C\in\mathfrak{C}_I: \ C\cap A\neq\phi \text{ and } C\cap A^c\neq\phi\}.$$

So, $\mathcal{P}_I(A)$ consists of those cells which intersect with *both* A and A^c, i.e., cells which are on the "perimeter" of A.

REGULARITY CONDITION (R.4): [Smoothness of Perimeter]

Denote $\mathfrak{F}_I := \{\overline{\Theta}, \overline{T}:\ T \in \mathfrak{T}_I\}$. We require

$$|I|^{\delta} \cdot \sup_{A \in \mathfrak{F}_I} \lambda(\mathcal{P}_I(A)) \to 0 \quad \text{as } |I| \to \infty.$$

Condition R.4 guarantees "smoothness" of the boundaries, relative to the grid: As the grid mesh becomes finer, the cell-wise approximation to the boundary (i.e., $\mathcal{P}_I(\cdot)$) must shrink in "area." This prohibits boundaries which wander through too many cells in $\mathcal{U}_d$ [see Example 1.c in Section 5].

It is trivial to directly check R.4 in the 1-dimensional case [see Section 5.1]. In the 2-dimensional case, we can actually reduce R.4 to a simple calculation of the *lengths* of boundaries. Consider a boundary in $\mathcal{U}_2$ that is expressible as a *rectifiable curve* r(t), i.e., r(·) is a continuous function from $[a, b] \subseteq \mathbf{R}^1$ into $\mathcal{U}_2$, with coordinates $r(t) = (r_1(t), r_2(t))$, satisfying

$$L(r) := \sup_{a = t_0 < t_1 < \ldots < t_k = b} \sum_{1 \le i \le k} \|r(t_i) - r(t_{i-1})\| < \infty.$$

The quantity L(r) is just the *length* of the boundary. The following relationship holds between the *perimeter cells* and the *length*:

THEOREM 3: Let d=2. If the set $A \in \mathfrak{F}_I$ corresponds to a boundary expressible as a *rectifiable curve* r(·), then

$$\lambda(\mathcal{P}_I(A)) \le 18 \cdot (L(r)+1)/\min\{n_1, n_2\}.$$

Proof of this result is in Appendix A.3. Now it is easy to check R.4 by approximating the lengths of boundaries [see Section 5.2].

5. APPLICATIONS

In this Section we explicitly apply the theoretical results of Section 4 to the Examples from Section 2. In each case a class of strongly consistent nonparametric boundary estimators is obtained. Note that R.1 is satisfied [by construction] in each of Examples 1.a, 1.b, 2.a, 2.b, 2.c; this regularity condition will not be further discussed.

5.1 The 1-Dimensional Case.

EXAMPLE 1.a: *The Change-Point Problem.* In R.2, take T_I to be the point $\lfloor \theta n_1 \rfloor / n_1$, where $\lfloor z \rfloor$ denotes the largest integer less than z. Since $\partial(\Theta, T_I) \leq 1/n_1$ and $|I| = n_1$, R.2 holds for any $\delta < \frac{1}{2}$. For R.3, we observe that $|\mathcal{T}_I| < |I|$; hence Proposition 3 applies. For each $A \in \mathfrak{F}_I$, there is only one "perimeter cell." Therefore $\lambda(\mathcal{P}_I(A)) = 1/n_1$, and R.4 is satisfied for any $\delta < \frac{1}{2}$. This establishes Theorems 1 & 2 for the class of change-point estimators $\hat{\Theta}_I$; in the rate factor, we can allow any $\delta < \frac{1}{2}$.

EXAMPLE 1.b: *The Epidemic-Change Model.* Take T_I to be the pair $\{(\lfloor \theta_1 n_1 \rfloor + 1)/n_1, \lfloor \theta_2 n_1 \rfloor / n_1\}$, so that T_I is an "inner approximation" to $\underline{\Theta}$. Again $\partial(\Theta, T_I)$ is of order $1/n_1$ and $|I| = n_1$, so R.2 holds for any $\delta < \frac{1}{2}$. Next we can apply Proposition 3, because $|\mathcal{T}_I| < |I|^2$. For $A \in \mathfrak{F}_I$ there are now two "perimeter cells," yielding $\lambda(\mathcal{P}_I(A)) = 2/n_1$ and satisfying R.4 for any $\delta < \frac{1}{2}$. Thus Theorems 1 & 2 hold: We have convergence of our nonparametric estimators for the epidemic-change model, with the rate factor allowing any $\delta < \frac{1}{2}$.

EXAMPLE 1.c: *Rationals vs. Irrationals.* We have heavily emphasized the set-theoretic nature of our approach -- in particular, we have exploited the partition $(\overline{\Theta}, \underline{\Theta})$ of $\mathcal{U}_d$. The sets $\overline{\Theta} := \{$*rational* numbers in $[0, 1]\}$ and $\underline{\Theta} := \{$*irrational* numbers in $[0, 1]\}$ constitute a perfectly legitimate partition of $\mathcal{U}_1$. Yet we would be surprised if the proposed method applied in this situation. Indeed, an obvious problem arises with the "smoothness" of the perimeter (R.4): *every* cell is in $\mathcal{P}_I(\overline{\Theta})$, so $\lambda(\mathcal{P}_I(\overline{\Theta})) \equiv 1$ and R.4 is violated.

5.2 The 2-Dimensional Case.

EXAMPLE 2.a: *Linear Bisection.* To check R.2, consider the following special case: Θ connects an endpoint on the *lower* edge of $\mathcal{U}_2$ to an endpoint on the *left* edge of $\mathcal{U}_2$; the coordinate on the lower edge is $\theta_1 \in (0, 1]$, and the coordinate on the left edge is $\theta_2 \in (0, 1]$.

Take T_I to be an analogously oriented segment, with lower edge coordinate $\lfloor\theta_1 n_1\rfloor/n_1$, and left edge coordinate $\lfloor\theta_2 n_2\rfloor/n_2$. Then $\partial(\Theta, T_I)=\frac{1}{2}(\theta_1\theta_2-\lfloor\theta_1 n_1\rfloor\lfloor\theta_2 n_2\rfloor/n_2 n_1)\leq 1/min\{n_1, n_2\}$, and hence R.2 is satisfied whenever

$$(\star) \qquad (max\{n_1, n_2\})^{\delta}(min\{n_1, n_2\})^{\delta-1} \to 0.$$

Other configurations of Θ similarly yield $(\star)$ as a sufficient condition for R.2. Since $|\mathfrak{T}_I|<|I|^2$, we can handle R.3 via Proposition 3. Now observe that $L(r)\leq\sqrt{2}$ for any $A\in\mathfrak{F}_I$, so by using Theorem 3 we find that R.4 is satisfied whenever $(\star)$ holds. Thus Theorems 1 & 2 apply to our nonparametric estimators of a linear bisecting boundary, provided $(\star)$ holds.

Condition $(\star)$ forces the grid design to asymptotically become finer in *both* dimensions. Consider in particular $min\{n_1, n_2\}=|I|^{\alpha}$ and $max\{n_1, n_2\}=|I|^{1-\alpha}$, where $0<\alpha\leq\frac{1}{2}$. Then $(\star)$ holds for any $\delta\in[0, \alpha)$. We obtain the best rate of convergence $|I|^{\delta}$ when we can afford the *symmetric* grid design, i.e., $\alpha=\frac{1}{2}$.

EXAMPLE 2.b: Templates. Consider the rectangular template Θ discussed in Section 2.2. Take T_I to be the candidate boundary corresponding to the largest $\underline{T}\subseteq\underline{\Theta}$. Then $\partial(\Theta, T_I)\leq\lambda(\underline{\Theta})-\lambda(\underline{T}_I)\leq 4/min\{n_1, n_2\}$, so that R.2 is satisfied whenever $(\star)$ holds. Since $|\mathfrak{T}_I|<|I|^2$, we can again use Proposition 3 to deal with R.3. Note that $L(r)\leq 4$ for all $A\in\mathfrak{F}_I$, so that Theorem 3 reduces R.4 to condition $(\star)$. Theorems 1 & 2 apply to our nonparametric estimators of rectangular templates, provided $(\star)$ holds. The analysis of the case $min\{n_1, n_2\}=|I|^{\alpha}$ and $max\{n_1, n_2\}=|I|^{1-\alpha}$, $0<\alpha\leq\frac{1}{2}$, is exactly as in Example 2.a.

EXAMPLE 2.c: Lipschitz Boundaries. For R.2, consider T_I corresponding to the piecewise linear function $y_{T_I}(\cdot)$ defined at each $z\in\{0/n_1, 1/n_1, \ldots, n_1/n_1\}$ by $y_{T_I}(z):=max\left\{i_2/n_2:\ i_2\in\{1, 2, \ldots, n_2-1\} \text{ and } i_2/n_2\leq y_{\Theta}(z')\ \ \forall z'\in[z-\frac{1}{n_1}, z+\frac{1}{n_1}]\right\}$. Note that $T_I\in\mathfrak{T}_I$, because for each $z\in\{1/n_1, 2/n_1, \ldots, n_1/n_1\}$ we have:

$$|y_{T_I}(z)-y_{T_I}(z-\tfrac{1}{n_1})|\leq sup_{z'\in[z-\frac{2}{n_1},\ z+\frac{1}{n_1}]}\, y_{\Theta}(z') - inf_{z'\in[z-\frac{2}{n_1},\ z+\frac{1}{n_1}]}\, y_{\Theta}(z') + 1/n_2\leq$$

$$\leq c_0\cdot 3/n_1 + 1/n_1.$$

Since $y_{T_I}(\cdot)$ is dominated by $y_\Theta(\cdot)$, we can write (for $|I|$ sufficiently large)

$$\partial(\Theta, T_I)= \sum_{z\in\{1/n_1,\, 2/n_1,\, \ldots,\, n_1/n_1\}} \int_{z-\frac{1}{n_1}}^{z} (y_\Theta(z')-y_{T_I}(z'))\, dz',$$

with each integrand bounded above by the r.h.s. of the preceding inequality. Therefore R.2 is satisfied whenever

$$(\dagger) \qquad (n_2)^\delta (n_1)^{\delta-1} \rightarrow 0.$$

For R.3, note that $|\mathfrak{T}_I| \approx (n_2)^{n_1}$, so Proposition 3 does not directly apply. In this situation, R.3 reduces to

$$(\dagger\dagger) \qquad \gamma(n_1 n_2)^{1-2\delta} - n_1 ln(n_2) \rightarrow \infty \quad \forall \gamma>0.$$

Lastly, we use Theorem 3 to handle R.4. Each $A\in\mathfrak{T}_I$ has $L(r)\leq\sqrt{c^2+1}$, because for $\overline{\Theta}$ we find:

$$\sum_{1\leq i\leq k} \|r(t_i)-r(t_{i-1})\| = \sum_{1\leq i\leq k} \sqrt{|y_\Theta(t_i)-y_\Theta(t_{i-1})|^2+|t_i-t_{i-1}|^2} \leq \sqrt{c^2+1}$$
$$\text{whenever } 0=t_0<t_1<\ldots<t_k=1,$$

and for $\overline{T}$ we find:

$$L(r)= \sum_{z\in\{1/n_1,\, 2/n_1,\, \ldots,\, n_1/n_1\}} \sqrt{|y_T(z)-y_T(z-\tfrac{1}{n_1})|^2+(\tfrac{1}{n_1})^2} \leq \sqrt{c^2+1}.$$

Thus R.4 reduces to ($\dagger$). Theorems 1 & 2 apply to our nonparametric estimators of Lipschitz boundaries, provided ($\dagger$) and ($\dagger\dagger$) are satisfied.

Consider the case $n_1=|I|^\alpha$ and $n_2=|I|^{1-\alpha}$, where $0<\alpha\leq\frac{1}{2}$. Conditions ($\dagger$) and ($\dagger\dagger$) will both be satisfied for any $\delta\in[0,\, min\{\alpha,\, \frac{1-\alpha}{2}\})$. Note that the best rate of convergence $|I|^\delta$ is obtained from the *non-symmetric* grid design with $\alpha=\frac{1}{3}$.

APPENDIX: PROOFS

A.1 Proof of Property 1.d in Proposition 1. Expand $\partial(\Theta, T)$ in terms of Θ, T, and T':

$$\partial(\Theta, T) = \min\{\lambda(\overline{T}\cap\underline{\Theta}\cap\overline{T}')+\lambda(\overline{T}\cap\underline{\Theta}\cap\underline{T}')+\lambda(\underline{T}\cap\overline{\Theta}\cap\overline{T}')+\lambda(\underline{T}\cap\overline{\Theta}\cap\underline{T}'),$$

$$\lambda(\underline{T}\cap\underline{\Theta}\cap\overline{T}')+\lambda(\underline{T}\cap\underline{\Theta}\cap\underline{T}')+\lambda(\overline{T}\cap\overline{\Theta}\cap\overline{T}')+\lambda(\overline{T}\cap\overline{\Theta}\cap\underline{T}')\} =: \min\{U, V\}.$$

Making the analogous expansions of $\partial(\Theta, T')$ and $\partial(T', T)$, in terms of Θ, T, and T', we see that $\partial(\Theta, T')+\partial(T', T)$ equals one of four possible expressions, e.g.,

$$W := \lambda(\overline{T}'\cap\underline{\Theta}\cap\overline{T})+\lambda(\overline{T}'\cap\underline{\Theta}\cap\underline{T})+\lambda(\underline{T}'\cap\overline{\Theta}\cap\overline{T})+\lambda(\underline{T}'\cap\overline{\Theta}\cap\underline{T})+$$

$$+\lambda(\overline{T}\cap\underline{T}'\cap\overline{\Theta})+\lambda(\overline{T}\cap\underline{T}'\cap\underline{\Theta})+\lambda(\underline{T}\cap\overline{T}'\cap\overline{\Theta})+\lambda(\underline{T}\cap\overline{T}'\cap\underline{\Theta}).$$

Since the summands in U are a subset of the summands in W, we have:

$$\partial(\Theta, T) = \min\{U, V\} \leq U \leq W.$$

Similar inequalities hold for the other three possible expressions of $\partial(\Theta, T')+\partial(T', T)$. □

A.2 Proof of Theorems 1 and 2. The main task is to establish:

(1) $\quad \mathbf{P}\{|I|^{\delta}\cdot\partial(\Theta, \hat{\Theta}_I)>\varepsilon\} \leq c_1\cdot|\mathcal{T}_I|\cdot exp\{-c_2\cdot\varepsilon^2\cdot|I|^{1-2\delta}\}$ for $|I|$ sufficiently large.

Then Theorem 2 is the case $\delta=0$. Theorem 1 follows from the Borel-Cantelli Lemma because, by R.3, $\sum_{|I|} |\mathcal{T}_I|\cdot exp\{-c_2\cdot\varepsilon^2\cdot|I|^{1-2\delta}\} \approx \sum_{|I|} exp\{-\frac{1}{2}\cdot c_2\cdot\varepsilon^2\cdot|I|^{1-2\delta}\} < \infty$. We shall obtain equation (1) via a series of Lemmas.

LEMMA 1: Define $\mathfrak{T}^*_I := \{\overline{\Theta}, \underline{\Theta}; \overline{T}, \underline{T}, \overline{T}\cap\overline{\Theta}, \underline{T}\cap\overline{\Theta}, \overline{T}\cap\underline{\Theta}, \underline{T}\cap\underline{\Theta}\colon T\in\mathcal{T}_I\}$. We have:

$$|I|^{\delta}\cdot sup_{A\in\mathfrak{T}^*_I} \left|\lambda(A)-|A|_I/|I|\right| \to 0 \quad \text{as} \quad |I|\to\infty.$$

Proof: Consider $A\in\mathfrak{T}^*_I$. Denote $\overline{\mathcal{P}_I}(A) := \{C\in\mathcal{C}_I\colon C\subseteq A\}\subseteq A$ and $\underline{\mathcal{P}_I}(A) := \{C\in\mathcal{C}_I\colon C\subseteq A^c\}\subseteq A^c$, so that $\{\mathcal{P}_I(A), \overline{\mathcal{P}_I}(A), \underline{\mathcal{P}_I}(A)\}$ form a partition of $\mathcal{C}_I$. There is a one-to-one correspondence between $\mathcal{C}_I$ and I, where each cell C is associated with a particular grid-node i which is a vertex of C. Now

$$\#\{C\in\overline{\mathcal{P}_I}(A)\}/|I| = \lambda(\overline{\mathcal{P}_I}(A)) \leq \lambda(A) =$$

$$= 1-\lambda(A^c) \leq 1-\lambda(\underline{\mathcal{P}_I}(A)) = [\#\{C\in\overline{\mathcal{P}_I}(A)\}+\#\{C\in\mathcal{P}_I(A)\}]/|I|,$$

and

$$\#\{C\in\overline{\mathcal{P}}_I(A)\} \le |\overline{\mathcal{P}}_I(A)|_I \le |A|_I =$$

$$= |I|-|A^c|_I \le |I|-|\underline{\mathcal{P}}_I(A)|_I \le |I|-\#\{C\in\underline{\mathcal{P}}_I(A)\} = \#\{C\in\overline{\mathcal{P}}_I(A)\}+\#\{C\in\mathcal{P}_I(A)\},$$

so that

$$\left|\lambda(A)-|A|_I/|I|\right| \le \left|\lambda(A)-\#\{C\in\overline{\mathcal{P}}_I(A)\}/|I|\right|+\left|\#\{C\in\overline{\mathcal{P}}_I(A)\}/|I|-|A|_I/|I|\right| \le$$

$$\le 2\cdot\#\{C\in\mathcal{P}_I(A)\}/|I| = 2\lambda(\mathcal{P}_I(A)).$$

Thus, sets A of the form $\overline{\Theta}$ and $\overline{T}$ are handled immediately by R.4. Sets A of the form $\underline{\Theta}$ and $\underline{T}$ are also handled by R.4, since $\mathcal{P}_I(A^c)=\mathcal{P}_I(A)$. For the remaining sets A in $\mathfrak{F}_I^*$, note that $\mathcal{P}_I(A_1\cap A_2)\subseteq\mathcal{P}_I(A_1)\cup\mathcal{P}_I(A_2)$, so R.4 applies again. □

We need some notation, making the dependence upon I explicit. The data will now be denoted $\{X_i^I : i\in I\}$, and analogously to Section 3 we denote:

$$\bar{h}_T^I(x):=\sum_{j\in\overline{T}} 1\{X_j^I\le x\}/|\overline{T}|_I \text{ and } \underline{h}_T^I(x):=\sum_{j\in\underline{T}} 1\{X_j^I\le x\}/|\underline{T}|_I,$$

$$d_{Ii}^T:=|\bar{h}_T^I(X_i^I)-\underline{h}_T^I(X_i^I)|,$$

$$D_I(T):=\lambda(\overline{T})\lambda(\underline{T})S_{|I|}(d_{Ii}^T:\ i\in I),$$

where $S_{|I|}(\cdot)$ is a mean-dominant norm with $|I|$ arguments. We also define:

$$\bar{\eta}_T(x):=[\lambda(\overline{T}\cap\overline{\Theta})F(x)+\lambda(\overline{T}\cap\underline{\Theta})G(x)]/\lambda(\overline{T}) \text{ and } \underline{\eta}_T(x):=[\lambda(\underline{T}\cap\overline{\Theta})F(x)+\lambda(\underline{T}\cap\underline{\Theta})G(x)]/\lambda(\underline{T}),$$

$$\delta_{Ii}^T:=|\bar{\eta}_T(X_i^I)-\underline{\eta}_T(X_i^I)|,$$

$$\Delta_I(T):=\lambda(\overline{T})\lambda(\underline{T})S_{|I|}(\delta_{Ii}^T:\ i\in I).$$

***LEMMA* 2:** For $|I|$ sufficiently large,

$$P\{|I|^\delta\cdot sup_{T\in\mathfrak{T}_I}|D_I(T)-\Delta_I(T)|>\varepsilon\} \le K\cdot|\mathfrak{T}_I|\cdot exp\{-k\cdot\varepsilon^2\cdot|I|^{1-2\delta}\}.$$

Proof: Denote $\overline{H}_T^{Ii}:=|\bar{h}_T^I(X_i^I)-\bar{\eta}_T(X_i^I)|$, $\underline{H}_T^{Ii}:=|\underline{h}_T^I(X_i^I)-\underline{\eta}_T(X_i^I)|$, and $e_{Ii}^T:=\overline{H}_T^{Ii}+\underline{H}_T^{Ii}$, so that $d_{Ii}^T\le e_{Ii}^T+\delta_{Ii}^T$ and $\delta_{Ii}^T\le e_{Ii}^T+d_{Ii}^T$. Then $D_I(T)-\Delta_I(T)\le\lambda(\overline{T})\lambda(\underline{T})S_{|I|}(e_{Ii}^T:\ i\in I)$, by virtue of properties (D.e) and (D.c) of $S_{|I|}$. The same bound applies to $\Delta_I(T)-D_I(T)$, yielding

$$|D_I(T)-\Delta_I(T)|\le\lambda(\overline{T})S_{|I|}(\overline{H}_T^{Ii}:\ i\in I) + \lambda(\underline{T})S_{|I|}(\underline{H}_T^{Ii}:\ i\in I),$$

by property (D.c).

Now observe that

$$\bar{H}_T^{Ii} \le \left|\frac{\sum_{j\in\bar{T}\cap\bar{\Theta}} I\{X_j^I \le X_i^I\}}{|\bar{T}\cap\bar{\Theta}|_I} - F(X_i^I)\right|\frac{|\bar{T}\cap\bar{\Theta}|_I}{|\bar{T}|_I} + F(X_i^I)\left|\frac{|\bar{T}\cap\bar{\Theta}|_I}{|\bar{T}|_I} - \frac{\lambda(\bar{T}\cap\bar{\Theta})}{\lambda(\bar{T})}\right|$$

$$+\left|\frac{\sum_{j\in\bar{T}\cap\underline{\Theta}} I\{X_j^I \le X_i^I\}}{|\bar{T}\cap\underline{\Theta}|_I} - G(X_i^I)\right|\frac{|\bar{T}\cap\underline{\Theta}|_I}{|\bar{T}|_I} + G(X_i^I)\left|\frac{|\bar{T}\cap\underline{\Theta}|_I}{|\bar{T}|_I} - \frac{\lambda(\bar{T}\cap\underline{\Theta})}{\lambda(\bar{T})}\right|.$$

The 1st modulus on the r.h.s. is bounded by $\bar{F}_T^I := sup_{x\in\mathbf{R}}\left|\frac{\sum_{j\in\bar{T}\cap\bar{\Theta}} I\{X_j^I \le x\}}{|\bar{T}\cap\bar{\Theta}|_I} - F(x)\right|$, and the factor $F(X_i^I)$ in the 2nd summand on the r.h.s. is bounded by 1. The 3rd and 4th summands on the r.h.s. are similarly bounded. Substituting these bounds into the r.h.s., we obtain $\bar{H}_T^{Ii} \le \bar{H}_T^I$ (say), where $\bar{H}_T^I$ does *not* depend upon *i*. Use an analogous argument to obtain $\underline{H}_T^{Ii} \le \underline{H}_T^I$. Therefore, applying properties (D.e), (D.b), and (D.d), we find:

(2) $$|D_I(T) - \Delta_I(T)| \le \lambda(\bar{T})\bar{H}_T^I + \lambda(\underline{T})\underline{H}_T^I.$$

The r.h.s. of equation (2) is comprised of four analogous summands. We shall deal explicitly with only the first one of these, i.e.,

$$\left|\left(\lambda(\bar{T}) - \frac{|\bar{T}|_I}{|I|}\right)\frac{|\bar{T}\cap\bar{\Theta}|_I}{|\bar{T}|_I} + \frac{|\bar{T}\cap\bar{\Theta}|_I}{|I|}\right|\bar{F}_T^I +$$

$$+\left|\left(\lambda(\bar{T}) - \frac{|\bar{T}|_I}{|I|}\right)\frac{|\bar{T}\cap\bar{\Theta}|_I}{|\bar{T}|_I} + \left(\frac{|\bar{T}\cap\bar{\Theta}|_I}{|I|} - \lambda(\bar{T}\cap\bar{\Theta})\right)\right|.$$

For the 2nd modulus, note that Lemma 1 applies -- uniformly in $T\in\mathcal{T}_I$. Also Lemma 1 applies to the 1st term inside the 1st modulus. It now suffices to consider

$$P\{sup_{T\in\mathcal{T}_I}(|\bar{T}\cap\bar{\Theta}|_I/|I|)\,\bar{F}_T^I > \varepsilon|I|^{-\delta}\} \le \sum_{T\in\mathcal{T}_I} P\{\bar{F}_T^I > \varepsilon|I|^{1-\delta}/|\bar{T}\cap\bar{\Theta}|_I\}.$$

Each probability in this summation is bounded by $K_1 exp\{-k_1\varepsilon^2|I|^{1-2\delta}\}$ (see Dvoretzky, Kiefer, & Wolfowitz (1956): their Lemma 2 and the discussion following their Theorem 3). □

LEMMA 3: We can write

$$\Delta_I(T) = \rho(T)\cdot S_{|I|}(\delta^{\Theta}_{Ii}:\ i\in I),$$

where $\rho(T) := |\lambda(\overline{T}\cap\overline{\Theta})\lambda(\underline{T}) - \lambda(\underline{T}\cap\overline{\Theta})\lambda(\overline{T})|$.

Proof: Observe that $\delta^{T}_{Ii} = \rho(T)\delta^{\Theta}_{Ii}/\lambda(\overline{T})\lambda(\underline{T})$. Now apply property (D.b) of $S_{|I|}$. □

LEMMA 4: For every $T\in\mathfrak{T}_I$, we have $\Delta_I(T) \leq \Delta_I(\Theta)$.

Proof: By Lemma 3, it suffices to show $\rho(T)\leq\rho(\Theta)$. In the definition of $\rho(T)$, consider the expression within the modulus. If this expression is positive, then $\rho(T)$ equals $\lambda(\overline{T}\cap\overline{\Theta})\lambda(\underline{T}\cap\underline{\Theta}) - \lambda(\underline{T}\cap\overline{\Theta})\lambda(\overline{T}\cap\underline{\Theta}) \leq \lambda(\overline{\Theta})\lambda(\underline{\Theta})$. The negative case is handled similarly. □

LEMMA 5: For $\gamma>0$,

$$\partial(\Theta, T)<\gamma \ \Rightarrow\ \rho(\Theta)-\rho(T)<\gamma$$

and

$$\partial(\Theta, T)>\gamma \ \Rightarrow\ \rho(\Theta)-\rho(T)>k'\cdot\gamma>0.$$

Proof: Note that

(3) $\quad \rho(\Theta)-\rho(T) = \min\{\lambda(\underline{\Theta}\cap\overline{T})\lambda(\overline{\Theta})+\lambda(\overline{\Theta}\cap\underline{T})\lambda(\underline{\Theta}),\ \lambda(\overline{\Theta}\cap\overline{T})\lambda(\underline{\Theta})+\lambda(\underline{\Theta}\cap\underline{T})\lambda(\overline{\Theta})\}.$

Comparing this expression to the definition of $\partial(\Theta, T)$, the first implication is clear. For the second implication, we have by hypothesis that $\max\{\lambda(\underline{\Theta}\cap\overline{T}),\ \lambda(\overline{\Theta}\cap\underline{T})\}>\frac{\gamma}{2}$ and that $\max\{\lambda(\overline{\Theta}\cap\overline{T}),\ \lambda(\underline{\Theta}\cap\underline{T})\}>\frac{\gamma}{2}$. Equation (3) then yields $\rho(\Theta)-\rho(T)>\min\{\lambda(\overline{\Theta}),\ \lambda(\underline{\Theta})\}\frac{\gamma}{2}$. □

LEMMA 6: For $|I|$ sufficiently large,

$$P\{|I|^{\delta}\cdot|\Delta_I(\hat{\Theta}_I)-\Delta_I(\Theta)|>\varepsilon\} \leq \widetilde{K}\cdot|\mathfrak{T}_I|\cdot exp\{-\widetilde{k}\cdot\varepsilon^2\cdot|I|^{1-2\delta}\}.$$

Proof: Let $T^0_I\in\mathfrak{T}_I$ be the maximizer of $\Delta_I(\cdot)$ over $\mathfrak{T}_I$. Then, by Lemma 4, we have $\Delta_I(\Theta)\geq\Delta_I(T^0_I)\geq\Delta_I(T)\ \forall T\in\mathfrak{T}_I$. And, by definition, we have $D_I(\hat{\Theta}_I)\geq D_I(T)\ \forall T\in\mathfrak{T}_I$. Now,

(4) $|\Delta_I(\hat{\Theta}_I)-\Delta_I(\Theta)| \leq |\Delta_I(\hat{\Theta}_I)-D_I(\hat{\Theta}_I)|+|D_I(\hat{\Theta}_I)-\Delta_I(T_I^0)|+|\Delta_I(T_I^0)-\Delta_I(\Theta)|$.

The second modulus on the r.h.s. is bounded by $sup_{T\in\mathcal{T}_I}|D_I(T)-\Delta_I(T)|$, because either $D_I(\hat{\Theta}_I)\geq\Delta_I(T_I^0)\geq\Delta_I(\hat{\Theta}_I)$ or $\Delta_I(T_I^0)\geq D_I(\hat{\Theta}_I)\geq D_I(T_I^0)$. The same bound applies to the first modulus on the r.h.s. of (4).

Using Lemmas 4 & 3, and properties (D.e) & (D.d) of $S_{|I|}$, the third modulus on the r.h.s. is bounded by: $\Delta_I(\Theta)-\Delta_I(T_I) = [\rho(\Theta)-\rho(T_I)]S_{|I|}(\delta_{Ii}^{\Theta}: i\in I) \leq \rho(\Theta)-\rho(T_I)$. For $|I|$ sufficiently large, R.2 ensures that $\partial(\Theta, T_I)<\frac{1}{2}\varepsilon|I|^{-\delta}$. Thus, by Lemma 5, the third modulus on the r.h.s. of equation (4) is *deterministically* bounded by $\frac{1}{2}\varepsilon|I|^{-\delta}$. Combining this with the bounds from the previous paragraph, we see that applying Lemma 2 completes this proof. □

__LEMMA 7:__ Denote

$$\mu_F:=\int_{-\infty}^{\infty}|F(x)-G(x)|dF(x),\quad \mu_G:=\int_{-\infty}^{\infty}|F(x)-G(x)|dG(x),\quad\text{and}\quad \mu:=\mu_F\lambda(\bar{\Theta})+\mu_G\lambda(\underline{\Theta}).$$

Then $\mu>0$.

<u>Proof:</u> By assumption we have $\Lambda:=\{x\in\mathbf{R}: |F(x)-G(x)|>0\}\neq\phi$. It suffices to show that either $\int_\Lambda dF(x)>0$ or $\int_\Lambda dG(x)>0$. The case where Λ contains a discontinuity point of F or G is trivial, so we will now presume that F and G are continuous at each $x\in\Lambda$.

Select $x_0\in\Lambda$ with (say) $F(x_0)>G(x_0)$. Then $\sigma:=\{y\in(-\infty, x_0): F(x)>G(x)\ \forall x\in(y, x_0]\}$ is nonempty, by continuity. Denote $y_0:=inf\{y\in\sigma\}$. If $y_0=-\infty$, then $(-\infty, x_0]\subseteq\Lambda$ and therefore $\int_\Lambda dF(x) \geq F(x_0) > G(x_0) \geq 0$. If $y_0>-\infty$, then $F(y_0)\leq G(y_0)$ and also $(y_0, x_0]\subseteq\Lambda$, yielding $\int_\Lambda dF(x) \geq F(x_0)-F(y_0) \geq F(x_0)-G(y_0) > G(x_0)-G(y_0) \geq 0$. □

__LEMMA 8:__ Equation (1) holds.

<u>Proof:</u> Applying Lemma 3, Lemma 4, Lemma 5, and property (D.f) of $S_{|I|}$, we have:

$$\partial(\Theta, T)>\gamma \;\Rightarrow\; |\Delta_I(\Theta)-\Delta_I(T)| = [\rho(\Theta)-\rho(T)]S_{|I|}(\delta_{Ii}^{\Theta}: i\in I) > k'\cdot\gamma\cdot\delta_I^{\Theta},$$

where $\delta_I^{\Theta}:=\sum_{i\in I}\delta_{Ii}^{\Theta}/|I|$. Thus,

$$P\{\partial(\Theta, \hat{\Theta}_I)>\varepsilon|I|^{-\delta}\} \leq P\{|\Delta_I(\Theta)-\Delta_I(\hat{\Theta}_I)|>k'\varepsilon|I|^{-\delta}\delta_I^{\Theta}\} \leq$$

$$\leq P\{|I|^{\delta}|\Delta_I(\Theta)-\Delta_I(\hat{\Theta}_I)|>k'\epsilon\omega\}+P\{\delta_I^{\Theta}<\omega\},$$

where $\omega:=\mu/2>0$ (by Lemma 7). The 1st probability on the r.h.s. is immediately handled by Lemma 6. The 2nd probability on the r.h.s. is bounded by $P\{|\delta_I^{\Theta}-\mu|>\omega\}$. Denote $\bar{\delta}_I^{\Theta}:=\sum_{i\in\bar{\Theta}}\delta_{Ii}^{\Theta}/|\bar{\Theta}|_I$ and $\underline{\delta}_I^{\Theta}:=\sum_{i\in\underline{\Theta}}\delta_{Ii}^{\Theta}/|\underline{\Theta}|_I$, so that $\delta_I^{\Theta}=\bar{\delta}_I^{\Theta}(|\bar{\Theta}|_I/|I|)+\underline{\delta}_I^{\Theta}(|\underline{\Theta}|_I/|I|)$. Notice that

$$|\delta_I^{\Theta}-\mu| \leq \bar{\delta}_I^{\Theta}\left|\frac{|\bar{\Theta}|_I}{|I|}-\lambda(\bar{\Theta})\right| + \lambda(\bar{\Theta})|\bar{\delta}_I^{\Theta}-\mu_F| + \underline{\delta}_I^{\Theta}\left|\frac{|\underline{\Theta}|_I}{|I|}-\lambda(\underline{\Theta})\right| + \lambda(\underline{\Theta})|\underline{\delta}_I^{\Theta}-\mu_G|.$$

The first and third summands on the r.h.s. are handled using Lemma 1. Consider the second summand on the r.h.s. (a similar argument holds for the fourth). By equation (2.3) of Hoeffding (1963), we have $P\{|\bar{\delta}_I^{\Theta}-\mu_F|>\omega/4\} \leq 2\cdot exp\{-\tilde{c}|\bar{\Theta}|_I\}$. Lemma 1 ensures that $|\bar{\Theta}|_I$ eventually exceeds $|I|\lambda(\bar{\Theta})/2$ (say); therefore the bound from Hoeffding (1963) can be absorbed into the earlier bound from Lemma 6, for $\epsilon>0$ sufficiently small. □

<u>A.3</u> <u>Proof of Theorem 3.</u> For any set $U\subseteq\mathfrak{U}_2$, let $\tilde{U}$ denote the *closure* of U, let $\mathfrak{B}(U):=\tilde{U}\cap\widetilde{U^c}$ denote the *boundary* of U, and let $\mathfrak{I}(U)$ denote the *interior* of U. For the set A we are assuming: $\mathfrak{B}(A)=\{r(t)\in\mathfrak{U}_2:\ t\in[a, b]\}$, with $\mathfrak{I}(A)\neq\phi$ and $\mathfrak{I}(A^c)\neq\phi$.

Define $K:=\lceil 2\cdot L(r)\cdot max\{n_1,\ n_2\}\rceil$. Consider intervals $[a_j,\ b_j]$, $1\leq j\leq K$, such that $a=a_1<b_1=a_2<b_2=\ \dots\ =a_K<b_K=b$. Let $r_j^*(t)$ be the curve $r(t)$ *restricted to* $t\in[a_j,\ b_j]$, and let $L_j:=L(r_j^*)$. In particular, choose the intervals so that $L_j=L(r)/K\ \forall j$. Denote $R_j:=\{r_j^*(t)\in\mathfrak{U}_2:\ t\in[a_j,\ b_j]\}$. Let $C_j\in\mathfrak{C}_I$ be a cell containing the point $r_j^*(a_j)$; and let $\mathfrak{C}_j$ be the collection of neighboring cells $C\in\mathfrak{C}_I$ that share a common edge or vertex with C_j (including C_j itself). Since $L_j\leq\frac{1}{2}(1/max\{n_1,\ n_2\})$, the points in R_j form a subset of the points in $\mathfrak{I}(\mathfrak{C}_j)$. Thus we have:

$$\#\{C\in\mathfrak{C}_I:\ C\in\mathfrak{P}_I(A)\} \leq \#\{C\in\mathfrak{C}_I:\ C\cap\mathfrak{B}(A)\neq\phi\} = \#\{C\in\mathfrak{C}_I:\ C\cap(\textstyle\bigcup_{1\leq j\leq K} R_j)\neq\phi\} =$$

$$= \textstyle\sum_{C\in\mathfrak{C}_I} \mathrm{I}\{\bigcup_{1\leq j\leq K} (R_j\cap C)\neq\phi\} \leq \sum_{1\leq j\leq K}\sum_{C\in\mathfrak{C}_I} \mathrm{I}\{R_j\cap C\neq\phi\} \leq$$

$$\leq \textstyle\sum_{1\leq j\leq K}\sum_{C\in\mathfrak{C}_I} \mathrm{I}\{C\in\mathfrak{C}_j\} \leq K\cdot 9.$$

Finally, recall that $\lambda(\mathfrak{P}_I(A)) = \#\{C\in\mathfrak{C}_I:\ C\in\mathfrak{P}_I(A)\}/n_1n_2$. □

REFERENCES

Asatryan, D. and Safaryan, I. (1986). Nonparametric methods for detecting changes in the properties of random sequences. In *Detection of Changes in Random Processes* (L. Telksnys, ed.), 1-13. Optimization Software, New York.

Bhattacharya, P. and Brockwell, P. (1976). The Minimum of an Additive Process with Applications to Signal Estimation and Storage Theory. *Z. Wahrsch. verw. Gebiete*, 37, 51-75.

Brodskii, B. and Darkhovskii, B. (1986). The a posteriori method of detecting the disruption of a random field. In *Detection of Changes in Random Processes* (L. Telksnys, ed.), 32-38. Optimization Software, New York.

Carlstein, E. (1988). Nonparametric change-point estimation. *Annals of Statistics*, 16, 188-197.

Csörgő, M. and Horváth, L. (1987). Nonparametric tests for the changepoint problem. *Journal of Statistical Planning and Inference*, 17, 1-9.

Darkhovskii, B. (1984). On two estimation problems for times of change of the probabilistic characteristics of a random sequence. *Theory of Probability and its Applications*, 29, 478-487.

Darkhovskii, B. (1985). A nonparametric method of estimating intervals of homogeneity for a random sequence. *Theory of Probability and its Applications*, 30, 845-849.

Darkhovskii, B. (1986). A general method for estimating the instant of change in the probabilistic characteristics of a random sequence. In *Detection of Changes in Random Processes* (L. Telksnys, ed.), 47-52. Optimization Software, New York.

Darkhovskii, B. and Brodskii, B. (1980). A posteriori detection of the "disorder" time of a random sequence. *Theory of Probability and its Applications*, 25, 624-628.

Deshayes, J. and Picard, D. (1981). Convergence de processus à double indice: application aux tests de rupture dans un modèle. *C.R. Acad. Sci. Paris*, 292, 449-452.

Dvoretzky, A., Kiefer, J., and Wolfowitz, J. (1956). Asymptotic Minimax Character of the Sample Distribution Function and of the Classical Multinomial Estimator. *Annals of Mathematical Statistics*, 27, 642-669.

Hoeffding, W. (1963). Probability Inequalities for Sums of Bounded Random Variables. *Journal of the American Statistical Association*, 58, 13-30.

Picard, D. (1985). Testing and estimating change-points in time series. *Advances in Applied Probability*, 17, 841-867.

Shaban, S. (1980). Change-Point Problem and Two-Phase Regression: An Annotated Bibliography. *International Statistical Review*, 48, 83-93.

Siegmund, D. (1986). Boundary Crossing Probabilities and Statistical Applications. *Annals of Statistics*, 14, 361-404.